PROPERTY OF
PRINCETON UNIVERSITY LIBRARY
RECEIVED FEB 26 1945

Coll. CG

STUDIES IN LINGUISTICS

George L. Trager
Editor

Martin Joos
Associate Editor

William Cornyn
Treasurer

Volume 2, Number 4 Winter, 1944

ARTICLES

H. M. Hoenigswald, Internal reconstruction 78

R. I. McDavid, Jr., Phonemic and semantic bifurcation: two examples 88

T. A. Sebeok, The Finno-Ugric language family 91

G. L. Trager, A projected Mongolian-English dictionary . . . 96

Each volume of STUDIES IN LINGUISTICS consists of four issues totalling about 100 pages. Subscription $1.50 Volume 2 is for 1943-44. Address editorial communications to George L. Trager, 1703 Harvard St., NW, Washington 9, D. C. Subscriptions should be sent to William Cornyn, Yale Graduate School, New Haven, Conn.

Published at
Washington, D. C.

STUDIES IN LINGUISTICS
Vol. 2, No. 4, 1944

INTERNAL RECONSTRUCTION

H. M. Hoenigswald, University of Pennsylvania

0. It is the purpose of this paper to discuss certain techniques which, while constantly used by scholars of comparative linguistics, do not seem to have ever been adequately described.

1.1 The term 'reconstruction' in linguistic work usually refers to the comparative method. The phonemes, morphemes, and constructions of two or more given languages presumed to be independent continuations of a common parent language are used in certain established ways to recover the phonemes, morphemes, and constructions of the proto-language. In some cases, of course, both the daughter languages and the proto-language (or a closely similar language) can be studied from records, as in the case of French, Portuguese, Rumanian vs. Proto-Romance, i.e. (Vulgar) Latin. In other cases, only the subsequent languages are given while the proto-language has to be reconstructed (Greek, Umbrian, Lithuanian, vs. Proto-IE). In still others, some or all of the daughter languages are already reconstructed proto-languages in themselves (such as Proto-IE and Hittite vs. Indo-Hittite, to the adherents of the Indo-Hittite hypothesis). The classical 19th century writers on IE comparative grammar have insisted that pedigree reconstruction of this kind yields, with greater or lesser accuracy, the linguistic structure of the parent language at the time of the cleavage. The greatest of them would not venture further back into what they called 'glottogonic speculations'. There have been, however, occasional lapses. Even the most conservative scholars, for instance, have held that there is some relationship between the IE accent and ablaut, and most of them would not hesitate to say that zero grade must have developed, at some time in IE history, from full grade under weak stress. The difference between this example and the more sweeping reconstructions of 'IE prehistorical grammar' by Hirt, Benveniste, Kuryłowicz, and others, is merely one of degree. On the whole, speculation without the benefit of comparative evidence has a reputation for being unsafe, a criticism which is, however, directed more often against certain far-reaching results than

against the procedure itself. The latter we shall call, for want of a better term, Internal Reconstruction.

1.2 More safely (if less spectacularly) than to proto-languages, the method of I[nternal] R[econstruction] is applied to ordinary linguistic material where it serves as a corollary or check for the observation of linguistic change in the records and for the results of comparative reconstruction. In turn, these more orthodox methods provide a gauge for the validity of IR. The following examples are, therefore, taken from those linguistic changes the complete history of which is known anyhow. In other words, we ask: How many of these changes could have been guessed if we had known nothing but their outcome?

2.11 Most dialects of colloquial modern German have voiceless versus voiced stops and spirants in contrast. The contrast, however, is not universal: it exists in Stunde : bunte, that is, in syllable initial, but not in und (i.e., /únt/) : bunt, that is in syllable final. This is merely a statement of phonemic distribution with no morphology involved. It may be found useful to symbolize this distribution as follows:

	'Total' range of positions
	Condition [A]=Syllable final
/t/	----------
/d/	----------

It is only as a result of this limitation, which is general in the language, that certain paradigms show /d/ before a vowel but /t/ at the end of a syllable. There is the pair Bunde : Bund (/búnde/ : /búnt/) with automatic alternation of this kind, while there are also pairs without such alternation, involving the more freely distributed phoneme, e.g. bunte : bunt (/búnte/ : /búnt/). This statement involves morphology; it is a morphophonemic statement. Starting from the diagram given above, we may symbolize such automatic alternation:

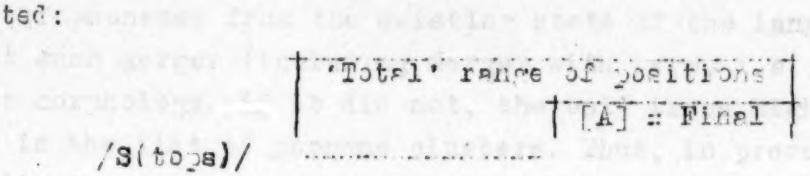
/d/ [-A] ~ /t/ [A]

It seems that IR, in such instances, invariably draws the conclusion

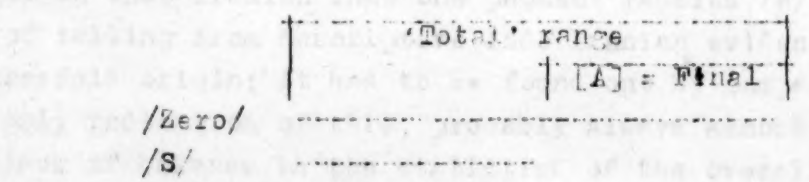
/d/ [A] > /t/ [A],

in other words, the phoneme of limited distribution was once free, and there has been conditioned sound change resulting in phonemic merger of /d/ and /t/, or 'neutralization' in position [A]. It appears essential, however, that in all pertinent cases morphological alternation in a paradigm be present. Here 'neutralization' does not prove anything. English /b/ and /p/ are also not in contrast after /s/, but this particular distribution is not the outcome of conditioned sound change.

2.12 The next example looks different on the surface, but is really of the same type. In ancient Greek final stops are not permitted:



This, again, is merely a statement of phonemic distribution. However, in inflectional paradigms, /S/ will alternate with Zero, e.g. /légon/ 'speaking' (nom. sg. neuter) ~ /légont-os/ (gen.); /méli/ 'honey' (nom.) ~ /mélit-os/ (gen.). Morphophonemically, then, the nominative forms are /légon0/ and /méli0/. Zero is obviously an element that has free distribution. As a matter of fact, its distribution is universal in the sense that any number of zeros may be set up between segmental phonemes. This may be meaningless and even misleading for other purposes, but it is useful here:



The historical interpretation is again

/S/ [A] > /Zero/ [A],

that is, stops used to have wider distribution; they also occurred finally (as is, of course, confirmed by the comparative method). There has intervened conditioned change resulting, once again, in the elimination of a contrast, in this case through loss of the phoneme in question, i.e., merger with Zero. Examples of this sort are extremely numerous. It is sufficient to mention one more case: the

limitation of English /k/ which is not permitted between pause and /n/. The alternation knowledge ~ acknowledge enables us to establish the sound change

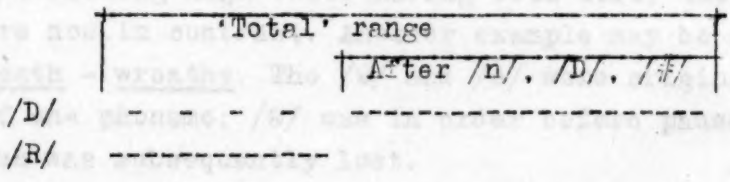
/k/ [#_n] > /zero/ [#_n]

2.13 The two cases examined so far exemplify the most accurate types of IR. It is comparatively simple to recover conditioned merger of phonemes from the existing state of the language provided that such merger (including merger with Zero, i.e., loss) involved some morphology. If it did not, the only trace might be a peculiar gap in the list of phoneme clusters. Thus, in present-day English (following a conservative phonemicization) /h/ is excluded from consonant clusters and position other than before stressed vowel. In Old English, however, its occurience was more regular: such clusters as the initial ones in hnecca 'neck' and hring existed, but were subsequently eliminated by loss of /h/. Since this did not entail morphological alternations, it leaves only an inconclusive trace in the lopsidedness of the occurrence of /h/.

2.2 Spontaneous elimination of contrasts, merger proper as well as loss, leaves no trace behind it. IE /e/, /o/, /a/ have all merged in Indo-Iranian into one phoneme labeled /a/. There is no way of telling from descriptive Indo-Iranian evidence that /a/ had a threefold origin; it had to be found out by comparative procedures. The only indication of this, probably always almost valueless, is the lack of balance in the statistics of the overall phonemic occurrence. In Sanskrit, /a/ is far more frequent than the other short vowels. Similarly, Modern Greek /i/ comes from nine different phonemic sources which have all merged unconditionally, and is consequently now the most frequent vowel. Further and better research into the statistics of phonemes may furnish some material in this connection. But it must be repeated that aside from this, the nature of the evidence will remain vague. Modern English would not enable us to recognize the different origin of heal and heel.[7] There is probably no indication in the structure of Old Irish of the unconditional loss of IE /p/ in Celtic.

2.31 The cases where increase in the number of phonemic con-

sound change may be either conditioned or spontaneous, increase of contrasts is, strictly speaking, always conditioned, the assumption of spontaneous split being inconsistent with the regularity of sound change. However, actual new contrast may be created 'spontaneously' as a result of foreign loans. A few trivial examples: In English we can set aside on a descriptive basis a certain part of the vocabulary as 'foreign'.[2] The phoneme /ž/ in initial position is found in that part of the lexicon only. This would make it possible to reconstruct the intrusion of /ž/ if an inference of this kind were necessary. The same is true of /v/. Here the contrast /f/ : /v/ was borrowed and made the old allophones /f/ : /v/ into phonemes. In Bengali and Hindustani a retroflex flap /R/ and a retroflex voiced stop /D/ are distributed almost complementarily: /D/ after /n/, itself, and /#/; /R/ in the other positions:



The diagram shows, however, that the distribution is not completely complementary. There is a thin overlapping on the part of the /D/. If we study the nature of the overlapping portion, we find that the occurrence of /D/ in abnormal positions is confined to foreign words like Hindustani /re:Di:o:/ 'radio'. All these data are purely distributional; there is no morphology involved.

2.32 Phonemic split, then, is mostly conditioned. The problem is whether it can be reconstructed from a subsequent stage of the language. It seems more convenient, in this case, to start out from the historical process rather than from the resulting structure. There seem to be two main types of phonemic split:

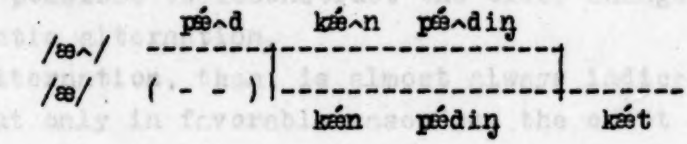
(1) The carry-over of an allophone, within a paradigm, into contrast. (It should be noted that this or any more elaborate formulation cannot pretend to account for the cause of the split. It is only a description of it.) An example is offered by the occurrence of /æ/ vs. /æ^/ in the variety of American English which has been called General Eastern.[3] The dialect alternates between /pǽ^d/ and /pǽdiŋ/ (noun), but has, in contrast with the latter, /pǽ^diŋ/ as

the participle of to pad; on the other hand, it says /kǽ^n/ 'tin container', but /kǽn/ 'am able'. Historically, /pǽ^d/ ~ /pǽdiŋ/ reflects the previous complementary distribution of what are now two phonemic entities, with /æ^/ before voiced consonant in syllable final vs. /æ/ in other positions. The form /kǽn/ is explained as restressed from weak position, where [æ] and not [æ^] prevailed.

(2) Sometimes the following situation arises: there are two allophones the distribution of which is determined by a phonemic feature somewhere in the utterance. Subsequently the determining phoneme is lost and the allophones enter into contrast. The vowel alternation in Modern Ger. Ofen : Öfen is now an alternation between two different phonemes. At an earlier stage of the language [o] and [ø] were allophones determined by a high vowel in a following syllable. The determining high vowel having been lost, the former allophones are now in contrast. Another example may be seen in cases like Eng. wreath ~ wreathe. The /θ/ and /ð/ were originally two allophones of one phoneme; /θ/ was in order before pause, /ð/ before a vowel, which was subsequently lost.

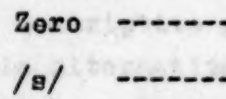
Incidentally, and purely as a matter of theory, it may be possible in a large number of cases to combine the two types listed above as being really one and the same process. Type (1), carry-over, may possibly be interpreted as a special case of (2), loss of determining feature, the latter being some sort of observable juncture at the morpheme boundary of /pǽ^d|iŋ/, or an observable special stress in the case of /kǽn/. This, however, is irrelevant for the present purpose. In both cases, the result is free morphological alternation (as against the automatic alternation in /búnt/ ~ /búnde/). The possibility of IR depends on how far the development has been carried. If it is still in the initial stage, the older structure can be recovered. The overall distribution of the new phonemes is often still peculiar. /θ/ and /ð/ are freely distributed now, but not quite as evenly as other consonant phonemes. Minimal contrasts exist but are hard to find (thigh : thy; wreath : wreathe). This would be even clearer if full statistical data were taken into account. But even so, it may be stated that /θ/ is frequent finally and initially, while /ð/ is frequent finally and medially between vowels (wreath, thumb : wreathe, weather) and rare or excluded in

most other positions. In General Eastern [æ^] is still excluded from position before /p t k ŋ l/, while [æ] is very rare in certain other positions:



In both these examples, then, the complementary pattern is still recognizable in the form of larger or smaller projecting portions in the diagram. Furthermore, in the favorable cases, the nature of the overlapping portion will be such that it contains morphologically related instances only. If so, all the morphologically isolated cases (not parts of a paradigm, such as /kǽt/ in the diagram), as well as some of the others, will be located in the projecting portions. In this way free alternation in a paradigm may be used to reconstruct phonemic split, provided that the structure has not been altered too much by other changes or by analogies.

3. It has been necessary to go to some length in this case because free alternation of phonemes in a paradigm may also develop from at least one other source. In Homeric Greek, for example, /s/ is of fairly free distribution as a consonant phoneme: /sê:ma/ 'monument', /pósis/ 'master', /stepanós/ 'wreath'. It is in free alternation, however, with Zero in cases like /génes-si/ vs. /géne-os/ 'kin' (two case forms). This instance of free alternation is different in two essential respects from the ones discussed above: (1) the projecting-overlapping pattern is missing, and (2) Zero is involved in the alternation. We know, of course, from the written records as well as from comparative evidence that the relationship was once automatic, due, in turn, to conditioned loss and subsequent limitation of the /s/ (see above):



In the course of time the alternation became free because a phone [s] was created from another source (e.g., [-si-] from [-ti-] by conditioned change as in /pósis/) and had to be put into the empty portion of the /s/ phoneme's distribution. Whenever Zero is a party

to free alternation it can be assumed that the relationship was, at a former period, automatic.[4] After this has been established it is of course possible to reconstruct the older change leading up to the automatic alternation.

Free alternation, then, is almost always indicative of previous change, but only in favorable cases can the exact character of the change be recovered through IR. Others may occasionally be decided on the basis of pattern analogy to clearer examples. For instance, the alternation /f/ : /v/ in present-day English is very obscure in character; the determining factor has disappeared, foreign influence has destroyed the remnants of the pattern, etc. But in some morphological categories it is obviously parallel to the pair /θ/ ~ /ð/ (knife : knives; mouth : mouths), which in turn shows clearer traces of the allophonic pattern. The linguist may thus assume an analogous history for /f/ ~ /v/.

4.0 The examples could be increased at will with interesting results. It is believed, however, that the foregoing types of procedure exhaust the major subdivisions in IR. The following is an attempt at summarizing them in two ways: according to the historical changes reconstructed, and according to the structural criteria used in the reconstruction.

4.1 The change must be truly phonemic, i.e., it must change the phonemic system rather than merely phonetic membership of the phonemes. Conditioned change can be recovered from phonemic alternations in morphology, provided that the system has not been obscured too much by other changes. Merger and loss of phonemes are more easily reconstructed than split. Spontaneous change can in favorable cases be reconstructed with the help of various considerations, involving morphological alternation of phonemes, overall pattern of phonemes, statistics, and the recognition of part of the vocabulary as 'foreign' on descriptive grounds.

4.2 Automatic alternation in a paradigm may prove merger or (if Zero is involved) loss through conditioned change. Free alternation in a paradigm may be used to prove split or, if Zero is involved, former automatic alternation implying previous loss. Quasi-complementary distribution of phonemes may be used to prove split or foreign intrusion. The nature of the overlapping portion in such a

pattern will help decide between the two possibilities. Pattern analogy between pairs of phonemes in free alternation may be used as an additional criterion.

5.1 These considerations are, and are meant to be, purely empirical. They are used implicitly in historical linguistic work, and are constantly confirmed by the more effective procedures of the linguistic interpretation of historically relevant records as well as by comparative grammar. In this they are not unlike the less vulnerable working rules of linguistic geography. The similarity goes even further: Both linguistic geography and IR use typical results of typical historical developments to reconstruct the developments where they are not known from other sources. Both must determine which events leave, as a rule, unambiguous traces (are reversible) and which do not. The validity of IR is restricted to the cases discussed above. It may also be limited to languages of a certain general structure (all our examples are IE and rely, perhaps, on the particular fashion in which phonemes and morphemes happen to function in IE), although there are indications that it is not. It does not of course admit of an exact chronology of the changes. The advantages of IR, on the other hand, are evident. Aside from being a useful instrument for the filling-in of gaps left by the orthodox methods, it is the only access to the prehistory of a language in the absence of other data, of a genealogically isolated language, or of a reconstructed proto-language with no known outside connection.

5.2 It is hoped that the method of Internal Reconstruction can be studied at a later time in greater detail. The reason for the utter neglect with which an important part of historical linguistic method has been treated is undoubtedly the lack of the proper descriptive terminology which alone permits one to formulate the scope and character of a phonetic change.[5] This terminology, however, has been provided for a considerable time now; not to make the widest possible use of it would mean to retard the progress of historical linguistics.

---

[1] It goes without saying that historical spelling, or conventional spelling based on morphophonemes, may give away a previous

sound structure. We are, however, concerned with evidence found in the spoken language only.

[2] See L. Bloomfield, Language 131, 153ff, etc. (New York, 1933).

[3] See G. L. Trager in American Speech 15.255-8 (1940).

[4] It would be meaningless, of course, to assume a development at the start of which Zero was an allophone of a phoneme. This makes the case discussed here different from the other example with free morphophonemic alternation.

[5] See, however, L. Bloomfield, Language[2], pp 369-391 passim, and, especially, R. Jakobson, TCLP 4.247-267, for a structural classification of sound changes. Jakobson does not use merger with Zero as the equivalent of loss.

STUDIES IN LINGUISTICS
Vol. 2, No. 4, 1944

PHONEMIC AND SEMANTIC BIFURCATION: TWO EXAMPLES

Raven I. McDavid, Jr., Southwestern Louisiana Institute

1. Dialects arise within a given speech-community when one of the phonemes of a phonetically related group of linguistic forms is replaced by another phoneme (or appears as an unfamiliar allophone) in the speech of one of the geographical or social groups of which the speech-community is composed. When the change affects a large group of words and a large section of a well-known speech-community, it may become known as a 'phonetic law'—such as those which bear the names of Grimm, Verner, and Sievers. Linguistically, these changes are of the same nature as the affrication of initial /t-/ in certain non-standard Danish dialects, or the similar affrication of /t-/ in various metropolitan dialects along the Middle Atlantic Seaboard.[1] The practical difference is that geographical and political cleavage made possible the persistence of the changes represented by the 'phonetic laws', while the peculiar features arising in any minor dialect today are likely to be lost under the impact of the more general dialects.

It is possible, however, that some of these divergences may persist in doublet-forms with a specialized standard usage and a phonemic shape somewhat different from that of the more generally-used word in the standard dialect. Such doublets are not uncommon in English: we have slick as well as sleek, parson as well as person, and the proper name Clark as well as clerk. It is the purpose of this article to call attention to two such doublets that are on the point of becoming established in the usage of the Southeastern United States.

2. Variations in the diphthongal syllabic of such words as join have been noticed at least since Pope rimed line and join. The phonetic drift of English during the past two centuries has been such that in almost all standard dialects the diphthongs of join and line are now quite distinct. In substandard dialects, to be sure, /ây/ forms of join, boil, etc., persist, but the /ôy/ forms have a far higher degree of social prestige. In most instances, the social cleavage between the two forms is clear, so that a transfer from a socially disapproved to a socially approved speech may be made—as

far as these words are concerned—by replacing /áy/ with the socially approved /óy/. The form hoist, however, is in a different category. In standard speech, the noun /hóyst/ refers to a mechanical contrivance, the verb to the execution of a mechanical or formalized operation, like the hoisting of a flag. In the vernacular—the speech of rural, substandard urban, or boys' groups—the noun /háyst/ refers to a lift or boost given with the arms or shoulders, the verb to the giving of such a lift. Since rural speakers hear of ammunition hoists /hóysts/, and socially privileged boys hear others ask to be given a /háyst/ up the side of a fence, the two forms are borrowed back and forth until for practical purposes they exist side by side as independent words.[2]

3. In the South Carolina Piedmont (as in most other dialects of present-day English) final or preconsonantal /-r/ may be preceded by two front vowels, /i/ and /æ/, the syllabics of ear and air respectively.[3] The distribution of various forms, and the free allophonic variations of these two phonemes, are somewhat different from what happens in other dialects, but the general pattern is the same. A more striking peculiarity is the existence of two forms—/rír/ and /rær/—for rear. The first of these is the general word; the second is a verb describing two types of action: (1) that of a horse rising on his hind legs, and (2) that of a man drawing himself back preparatory to throwing a missile or striking a blow, as 'I /rærd/ back and hit him', or 'He /rærd/ back and threw the ball as hard as he could.' Whether decreasing familiarity with horses will make the metaphor less apparent and keep the doublet from spreading is a matter for speculation. The present existence of the doublet in at least one dialect is a fact.

4. These examples illustrate the persistence of one of the processes by which the vocabulary of a language may be enriched. If two groups of people in the same speech-community use minimally different forms to describe the same action, state, or thing, it is not necessary that one form shall drive out the other; both forms may remain, either or both with specialization of meaning.

[October 17, 1944]

---

[1] This affrication was first noticed by Henry Lee Smith, Jr., in

his studies of contemporary standard American dialects.

[2] The specialization of the two forms was recognized by rural informants interviewed for the Linguistic Atlas in Upper South Carolina during the summer of 1941. Speakers from other dialect-areas, urban as well as rural, have since confirmed this differentiation.

Roil, of course, has no /óy/ form in normal speech, but is almost universally /ráyl/. It is fairly common in speech, but appears in writing so infrequently that there can hardly be said to be a written form to which exponents of correct pronunciation may appeal. [Those who know the word only from its written form, as I do, will say /róyl/ on the rare occasions when they see it.—GLT] [Because roil is a book-word in Wisconsin, we purists never used it to replace rile; instead we used muddy or stir up in speaking of water in a spring, stir up or make mad in speaking of a person. Those of us who learn the word roil use it in the first sense only, and few if any are aware that it is the book-dialect equivalent of rile. This is what makes the situation of rile : roil different from that histe : hoist; I don't suppose there are many people left today who don't know the word hoist as the 'correct' pronunciation of histe. Practically the same is the situation of sassy : saucy: like rile, sassy is a homely word, though not quite as many purists avoid it as avoid rile, perhaps because they have no convenient equivalent like stir up for rile, but would have to use some such affected word as impertinent. And the book-word saucy is usually learned, if it is learned at all, as an independent word, without discovering any connection between it and sassy. (Sass for the noun sauce is obsolete; the verb sass has the same status as rile and is similarly replaced by answer back, with no temptation to adopt sauce instead because the literary language no longer recognizes a verb sauce.)

[We note that histe and sass can both be spelled. In the case of histe this fact is of little moment, since the universal awareness that it is a pronunciation of hoist makes the spelling irrelevant. In the case of sass it is of great moment, as can be seen from what has been said above.

[Different from both these cases is that of the word teat. The native pronunciation rimes it with bit. But anybody who wrote tit would be told that the word is properly written teat. The notion that it ought to rime with beet is widespread in rural areas, but it conflicts with a feeling that such a pronunciation would be affected. This conflict cooperates with a feeling of delicacy to eliminate the word entirely from the speech of many persons, who thus become unable to talk about the object at all, since there is no agreed-upon equivalent name.—MJ]

[3] The first of these varies rather widely, from mid-high to mid-mid, so that G. L. Trager suggests that /e/ might be a better symbol than /i/. Whichever symbol is chosen, the range of variation is the same; the pertinent statement is that there is no phonemic contrast between /i/ and /e/ before /-r/, but one phoneme with a wide range of free variation.

STUDIES IN LINGUISTICS
Volume 2, No. 4, 1944

THE FINNO-UGRIC LANGUAGE FAMILY

Thomas A. Sebeok, Indiana University

[This paper is an outline or summary of a lecture before the Language Club of Indiana University on November 7, 1944; as such, it necessarily lacks examples and other scholarly compendia.]

11. Before discussing the Finno-Ugric language family as such, we must come to an understanding as to what constitutes a linguistic family or, more generally, what relationships can there exist between two given languages, A and B?

12. A and B have certain universal features in common. All languages, hence also A and B, can be analyzed into phonemes, including both consonants and vowels, morphemes, words, and so forth. These features are obviously of no use in a classification of languages.

13. A and B have certain accidental features in common. This is statistically probable, as shown by Shirokogoroff (Uralic and Tunguz) and Lewy (Uralic and Kičé). For example, Hungarian has a word fiu 'boy' and Rumanian has a word fiu 'boy'. But while the Hungarian word is known to derive from an older form of Hungarian, pi plus a diminutive suffix -u (-p), the Rumanian form comes from the Latin filius. Similarly, the Hungarian word év 'year' and the Rumanian ev 'period of time, age' have different histories: the Hungarian word is to be compared with Finnish ikä, Cheremiss i, Mordvin ije, Lapp jâkke, while the Rumanian comes from the Latin aevum.

14. A and B have a large number of typical structural features in common. We say that they belong to the same 'linguistic area'. This is Velten's translation of the German term Sprachbund, the French alliance linguistique. For example, the Russian domain, including White Russian and Ukrainian, and Polish, are opposed to the Czecho-Slovak region by, among other things, their lack of quantity in the vowel system; they share this feature with the Finno-Ugric and Turkic languages of European Russia, while other languages of the Finno-Ugric family (e.g. Hungarian), and some Turkic languages, possess this feature, and from this point of view belong to the same linguistic area as Czecho-Slovak. Likewise, Estonian belongs to the same linguistic area as certain Germanic languages (e.g. Swedish), from the point of view of tone. [But Estonian 'tone' is really a matter of quantity.—GLT]

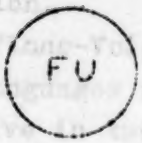
15. A and B are later forms of a single earlier language, C. We say that A, B, C, belong to the same language family. Thus, Spanish, Italian, Rumanian, French, Portuguese are later forms of a single earlier language, Latin.

21. Assume the existence of such a language C, called Uralic, in the Northeastern part of European Russia, sometime around the 3rd millenium B. C. By certain linguistic techniques, we can reconstruct some of the material culture, social organization, and daily life of this people. Thus, we know that they had a gathering—a hunting-fishing—economy. We believe that the dog, the reindeer, and possibly the sheep, were their only domestic animals. They lived in conical, skin-covered tents, open at the top, and during the winter moved into subterranean huts. They lived an extended family life as shown by a rich kinship terminology; we know almost nothing of their political organization. Their religion is presumed to have been animistic, including a cult of the dead, worship of the bear and other animals, and magic practices.

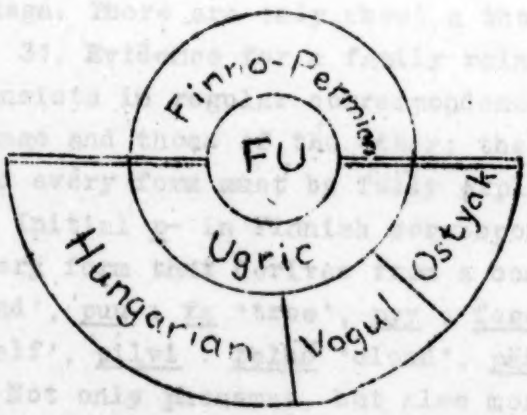
22. Next, we must assume that, for some reason—say, lack of sufficient food supply—a part of the tribe split away, emigrated into the northeast of Siberia. Living in isolation from its relatives, but exposed to a set of new (so-called circumpolar) influences, this people developed a walrus and bear hunting economy. It now speaks five fairly divergent dialects, known collectively as Samoyed (in Russian, Nenets).

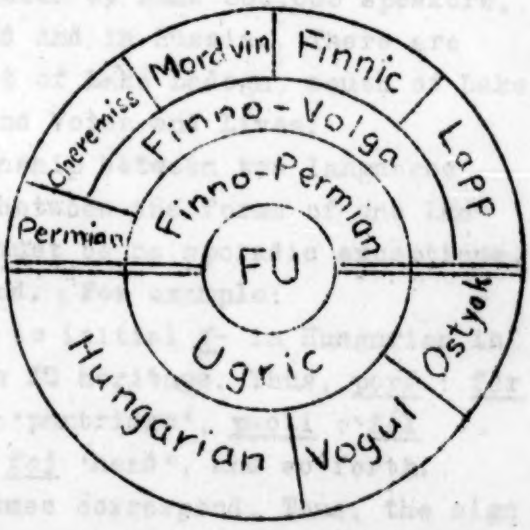
23. Meanwhile, the remainder of the group develops along lines of its own. It comes into contact (commercial; selling squirrel skins, for instance) with speakers of Indo-European, as evidenced by such words as Finnish <u>myy-</u> 'to sell', <u>osta-</u> 'to buy', Hungarian <u>viz</u> 'water', <u>vezet</u> 'he leads', and somewhat later with speakers of Indo-Iranian dialects (Hungarian <u>száz</u>, Finnish <u>sata</u> 'hundred', Hungarian <u>arany</u> 'gold', <u>méh</u> 'bee', <u>méz</u> 'honey', <u>szarv</u> 'horn', etc.) The group is known as Finno-Ugric (FU); it may be depicted by a circle (see page 93).

24. Again, the FU group expands in several directions. The first break-up resulted in two divisions, including languages known as Finno-Permian and Ugric:



25. Ugric developed into three languages, spoken today. The first of these is Ostyak (Khanti). Ostyak is spoken along the lower reaches of the river Ob, by some 18,000 speakers. The second is Vogul (Mantsi), spoken along the banks of the rivers Sosva, Sygva, Lozva, Vagilsk, Pelymka, Konda, and Tavda, by some 5000 speakers. The third of the Ugric languages is spoken in Hungary. The Hungarians entered the Carpathian basin in the 9th century. Previously, they were located east of the Ostyaks. They came into early contact with Turkic tribes; this transformed their culture as well as language. After a long period of migration, during which they came into contact with many peoples speaking Turkic and Indo-European languages, they settled in their present surroundings.





26. Meanwhile, Finno-Permian also split into two groups, Permian and Finno-Volga. There are two Permian languages spoken today: Ziryen (Komi) and Votyak (Udmurt). The Ziryens are located especially in the northeastern corner of European Russia; there are about

364,000. Their writing dates from the 14th century. The Votyaks live mostly between the rivers Vyatka and Kama, numbering about half a million.

27. The Finno-Volga group splits in four directions. Of these, the Volga languages are Cheremiss (Mari) and Mordvin (Mordva). The Cheremiss live in two big groups to the northeast and to the northwest of the Volga bend, numbering about 425,000. The Mordvins live to the south of the Cheremiss and the Votyaks, and there are about one and a quarter million of them.

28. In the northernmost part of Norway, Sweden, Finland, and Russia live about 33,000 speakers of Lapp. The Lapps are one of the historical mysteries of the FU group. Lapp is certainly a FU language, though it shows peculiar characteristics not shared by any other.

29. The Finnic group includes the following modern languages: Finnish (or Suomi), Estonian, Karelian, Vepse, Vote, and Live. The group, before splitting up into the languages mentioned, underwent a strong influence from speakers of Baltic languages, then of some Germanic languages, and also Slavic. The Finns occupy Finland, the Estonians Estonia and parts of Latvia. Karelian, which is linguistically very close to Finnish, is spoken by some 300,000 speakers, in the southeastern corner of Finland and in Russia. There are some 30,000 Vepse, in the region east of Lake Ladoga, south of Lake Onega. There are only about a thousand Votes and Lives.

31. Evidence for a family relationship between two languages consists in regular correspondences between the forms of one language and those of the other; there must be no sporadic exceptions, and every form must be fully explained. For example:

Initial p- in Finnish corresponds to initial f- in Hungarian in every form that derives from a common FU heritage. Thus, perä : far 'end', puu : fa 'tree', pyy : fogoly 'partridge', puoli : fél 'half', pilvi : felhő 'cloud', pää : fej 'head', and so forth.

Not only phonemes, but also morphemes correspond. Thus, the sign of the plural is -k in Hungarian, -k in Lapp: Hungarian halak 'fishes', Lapp guolek 'fishes'; Hungarian szavak 'words', Lapp sānek 'words', Hungarian hallanak 'they hear', Lapp gullik 'they hear', etc.

Final -(b)b in Hungarian and Ziryen corresponds to final -mp in

Finnish and Lapp. The suffix indicates the comparative degree of adjectives: Hungarian nagyobb 'bigger', Finnish suuremp-, Lapp šoreamp 'bigger', etc.

41. Certain structural similarities—e.g., vocalic harmony—and a number of common etymologies, have led some workers to relate the Uralic family, without the careful linguistic process just described, to the Altaic group of languages (including Turkic, Mongol, and Tunguz). The thus formed Ural-Altaic group may turn out to be a linguistic area rather than a linguistic family; in any case, the relationship remains to be proved. Uralic has also be related, without too much success on the whole, to Yukaghir (by Collinder), Eskimo (by Sauvegeot), Japanese (by Winkler), Indo-European (by Jakobsohn, etc.), Munda (by Hevesy), Dravidian (by Burrow), and so forth.

51. Among the work that remains to be done in the Finno-Ugric field, the following points should be emphasized.

a) Preparation of structural grammars, especially of certain hitherto poorly described languages of the Russian area.

b) Basing of all structural material either on work in the field or on aid of imported native informants.

c) Reconstruction by modern structural-historical techniques to clear up many problems.

d) A careful survey and comparative study of dialects within each language with a view to illuminating problems of the relationship of language and area.

STUDIES IN LINGUISTICS
Vol. 2, No. 4, 1944

A PROJECTED MONGOLIAN-ENGLISH DICTIONARY

George L. Trager, Washington, D. C.

In the course of its work in recommending forms of geographical names for the use of other government agencies, the United States Board on Geographical Names has had occasion to deal with maps and other materials containing large numbers of names in Mongolia (also known as Outer Mongolia, and as the Mongolian People's Republic). The sources containing these names are very uneven in linguistic reliability, and the problems of determining the best form of a Mongolian name are numerous and difficult.

It became my duty as Chief of the Linguistic Section of the Board to study the materials and develop means of processing the sources so that consistent and reliable results might be achieved. Examination of available studies indicated that all Mongolian place names are words or phrases which have meaning in the spoken language of today. It was accordingly decided to establish a procedure whereby the names found in the map sources are transformed into a standardized Mongolian orthography. Thus the name of a lake which appears on Russian-made maps as Xara Nur (Хара Нур), and in various other sources with the first element often spelled Kara and the second Nor, comes out in the normalized form as Hara Nuur (hara 'black', nuur 'lake'). This form was determined from glossaries and other recent linguistic materials of Russian origin. The method of processing is described, with bibliographical references, in the Board's forthcoming Special Publication No. 50, Guide to Geographical Names in Tibet, Sinkiang, and Mongolia.

As a result of the method evolved, and of the need for editing a large number of maps containing thousands of Mongolian names, it was decided to prepare a complete Mongolian-English index to the approximately 15,000 entries in A. V. Burdukov's Russian-Mongolian dictionary, Leningrad, 1935. This index, kept on file at the Board's offices, becomes the official 'best source' for Mongolian placenames to be processed by the method described in the Guide. In the absence of the possibility of first-hand checking, Burdukov's work is, of course, subject to correction, but its internal consistency is a good indication of its general accuracy.

With the preparation of the index, the possibility arose of working it up into a useful dictionary. Burdukov's dictionary contains no examples, and where a Russian word is glossed by more than one Mongolian word, or a Mongolian word is given as the equivalent of several Russian words, there seems at first to be no way to determine the appropriate meanings to be used in English translation. However, the dictionary has a brief grammatical introduction by N. N. Poppe (p. 7-43), and Poppe has also written a 'Practical textbook of the Mongolian spoken language (Halha dialect)' [Praktičeskij učebnik mongolʼskogo razgovornogo ăzyka (xalxaskoe narečie)], Leningrad, 1931.

By means of the rather transparent grammatical analysis of many of the entries, and by comparison of multiple glosses, it is possible to work up a fairly reliable translation for most of the lexicon. When this became apparent, I recommended that the preparation of the index as a dictionary, and its eventual publication, be undertaken as a project which, while directly aiding in the specific functions of the Board, would also be of considerable general value. The project has been approved, and the preparation of the material is now going on.

An enlarged photostate copy of the Burdukov dictionary was cut up and the entries pasted on 3 x 5 cards. A retranscription of the Mongolian gloss into the standardized orthography determined as best for geographical names was entered in the upper left of the card. These cards, alphabetically filed, constitute the index. The addition of English translations and their collation and final editing constitutes the work now in process.

Burdukov used an essentially (but in my opinion not entirely) correct phonemic analysis of Halha Mongolian, and wrote the language in the Latin alphabet devised for Buryat-Mongolian. The symbols were: a, b, c [ts], ç [č], d, e, g, i, j [y], k [x], l, m, n, o, e [ŏ], p, r, s, ş [š], t, u, y [ü], z [dz], ƶ [ǰ]; f, h, v were used in loanwords; long vowels were written double— aa, ee, etc., and diphthongs as ai, ei, etc. This orthography was apparently used in Mongolia for some time, but was later replaced, it would appear, by a cyrillic orthography. I have seen maps and indexes in this latter, and reconstruct it as follows (in the order of the equivalents of Burdukov's list above):

а, б, ц, ч, д, э, г, и, й (also я, е, ьи and йи, ю for Burdukov's ja, je, ji, ju), х, л, м, н, о, ӧ, п, р, с, ш, т, ў (for Burdukov's u), у (for Burdukov's y), цз, чж; long vowels as doubles or as ā, ē, etc.; diphthongs as ай, эй, etc. The Board uses this cyrillic alphabet as the basis of its standardized spelling, thus (in the same order again): a, b, ts, ch, d, e, g, i, y, h, l, m, n, o, ö, p, r, s, sh, t, u, ü, dz, j; long vowels as doubles, aa, ee, etc.; diphthongs as ay, ey, etc. This orthography will be used in the dictionary in order to make the latter most useful and usable although it is recognized that c, č, š, z (instead of ts, ch, sh, dz) would be more desirable on a purely linguistic basis (thus, the sequence s+h has to be written s.h to distinguish it from sh, while if š were used, the special use of the dot would be unnecessary).[1]

The dictionary will be accompanied by a brief grammatical sketch constituting a reworking of Poppe's material in the light of somewhat stricter methods of linguistic analysis than those used by Russian linguists, and taking into account other available studies. I hope to publish in SIL in the near future a somewhat more detailed version (especially in phonology) of this grammatical sketch.

---

[1] After this article had been approved for publication and the stencils had been made, it was decided to use kh instead of h; this decision was made for reasons that have no linguistic validity, and though concurred in by the writer does not seem to him to be either necessary or desirable.